



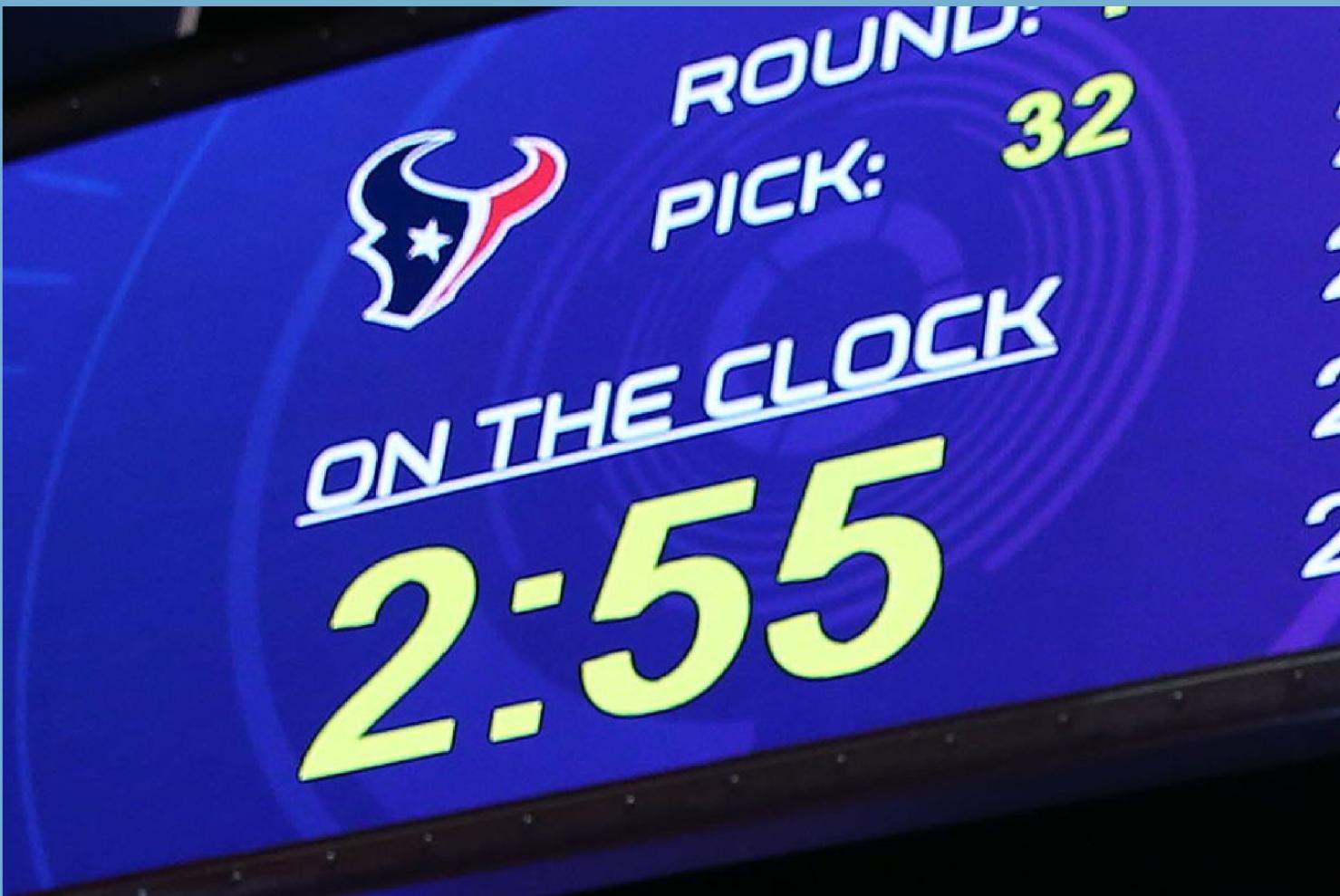
THE NFL DRAFT

Looking at Scouting Reports, the Combine,
and College Performance



THE PROBLEM

- How can the NFL teams successfully predict when players will get taken in the NFL draft
 - When to make a trade
 - Tie-Breaker
- How important is the combine versus the scouting reports vs their college performance





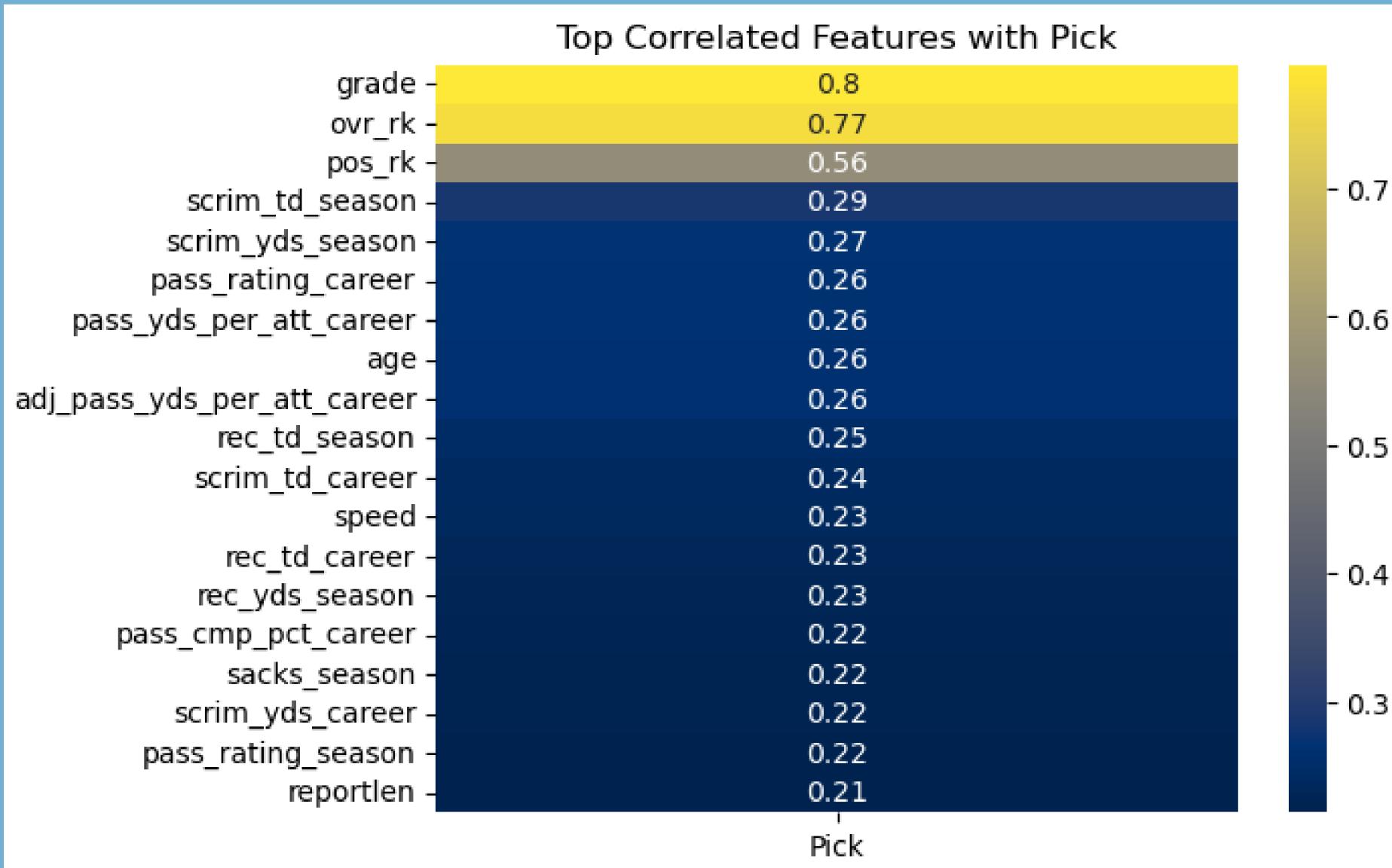
THE DATA

- Draft information, combine testing, and college stats web scraped from sportsreference.com
 - Scouting reports were taken from ESPN
 - Every combine and draft since 2005



SUMMARY

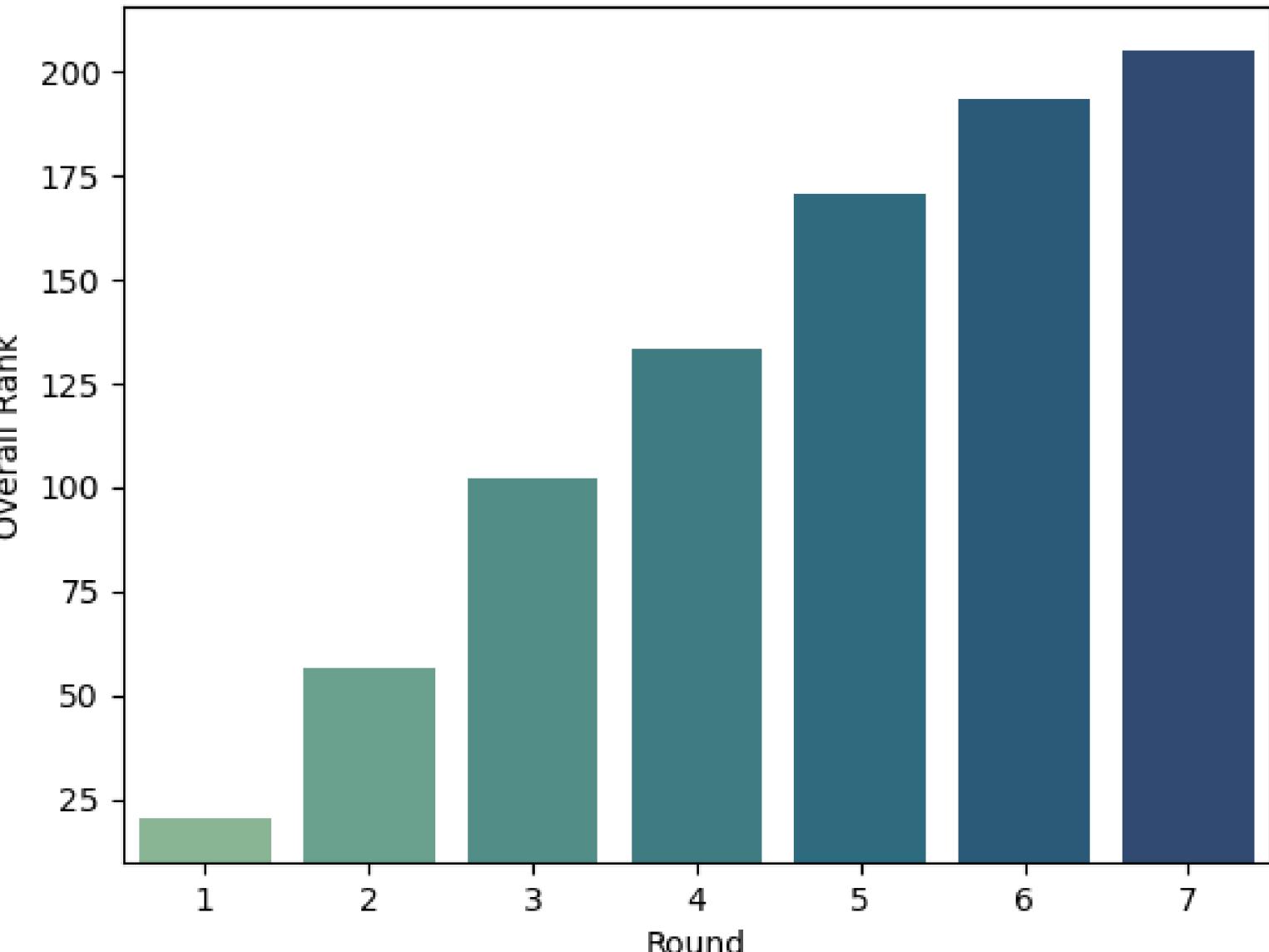
- Grade / Overall Rank are extremely significant in the determining where a player will get pick
- Length of the scouting report, combine metrics, and age make up some of the better lower tier features
- Offensive stats are more significant than defensive stats



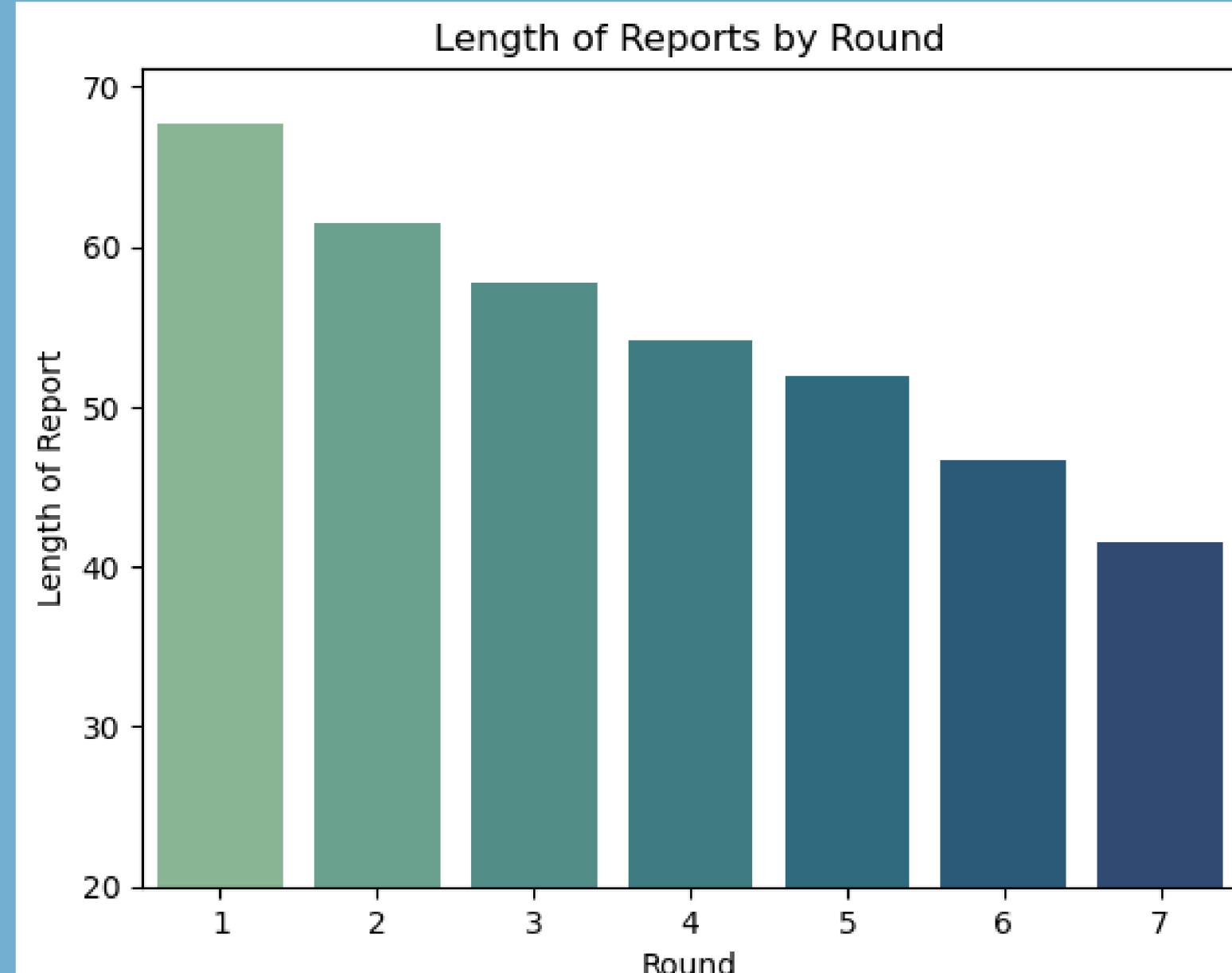


ROUND BREAKDOWN

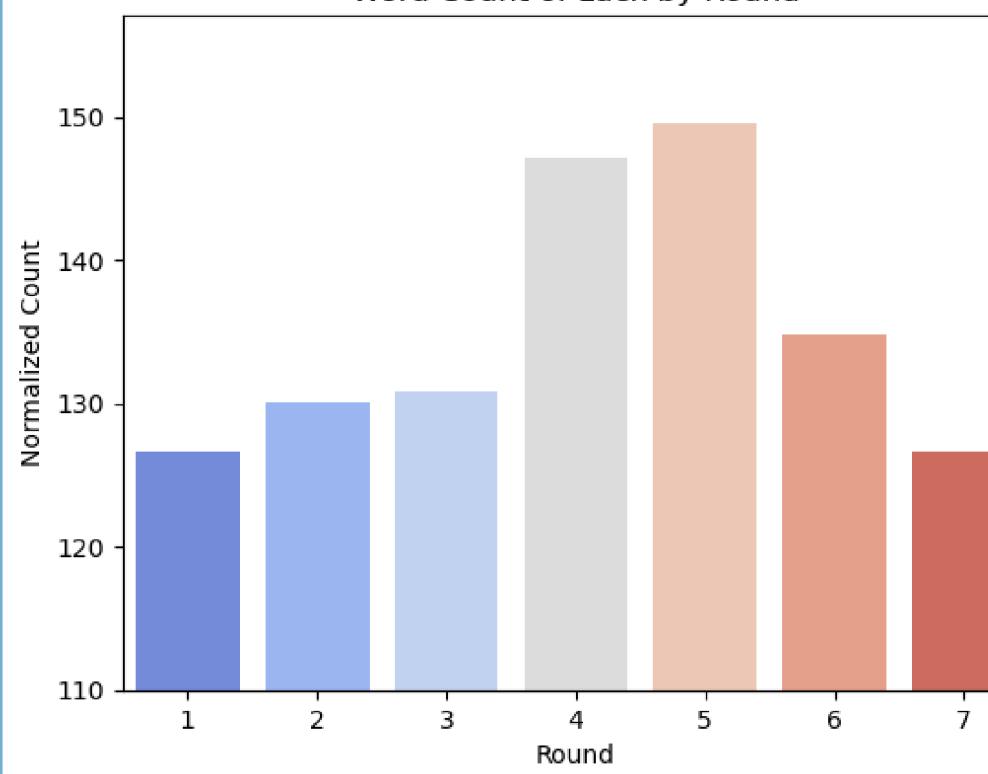
Overall Rank by Round



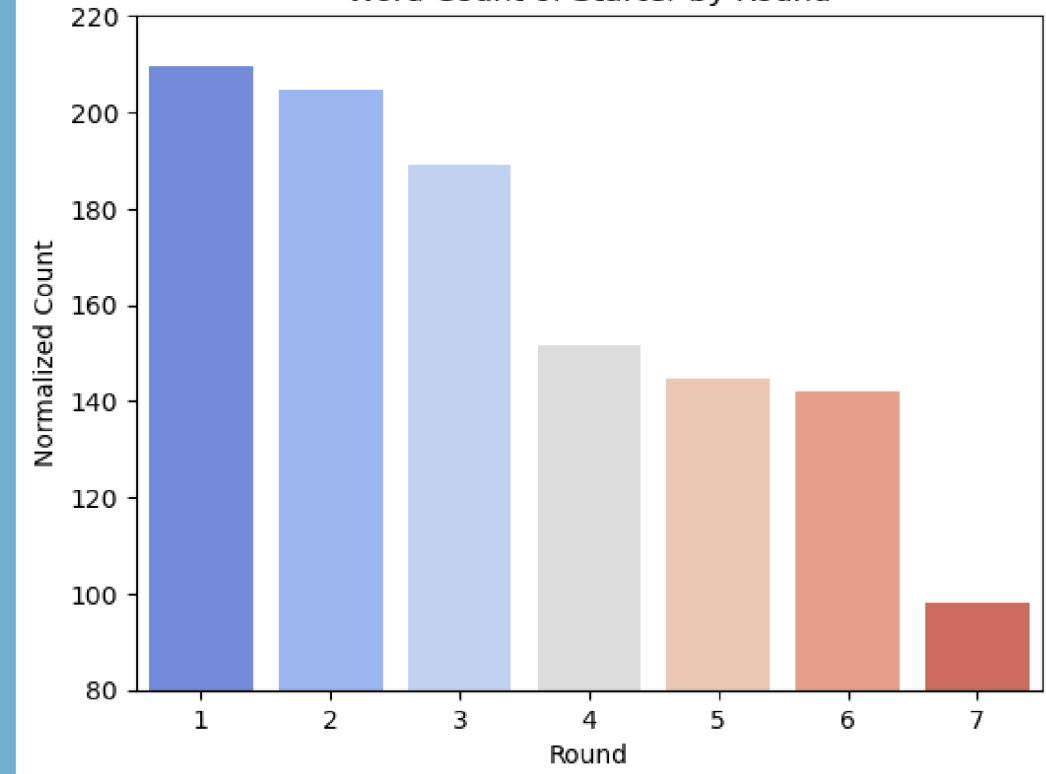
Length of Reports by Round



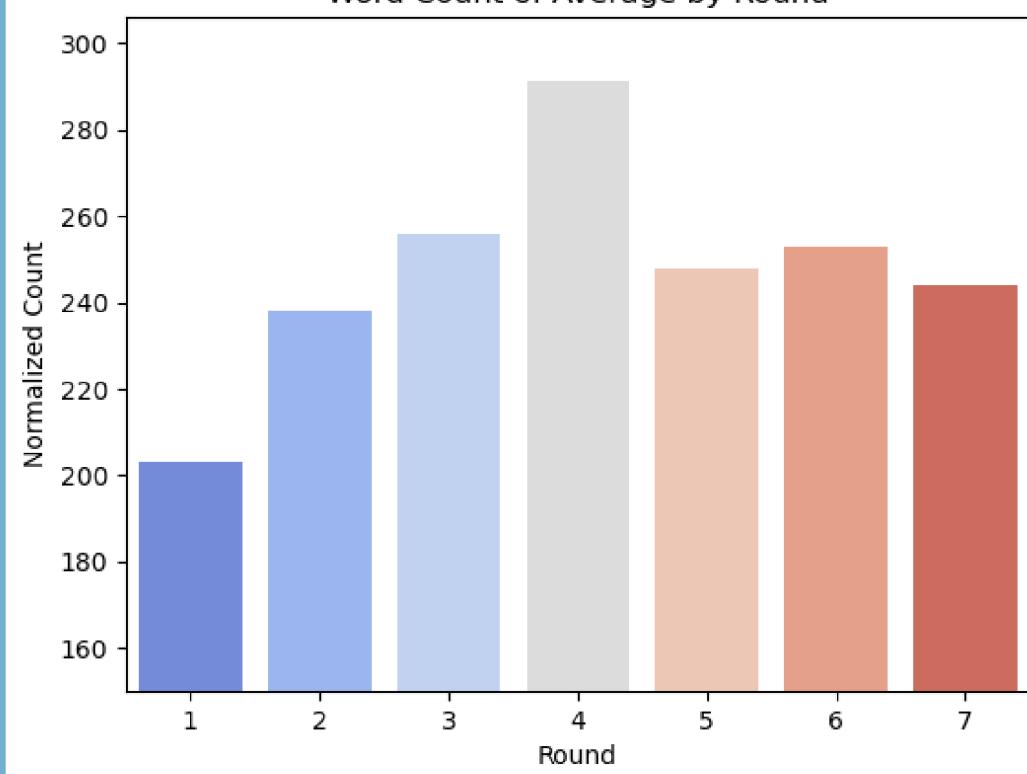
Word Count of Lack by Round



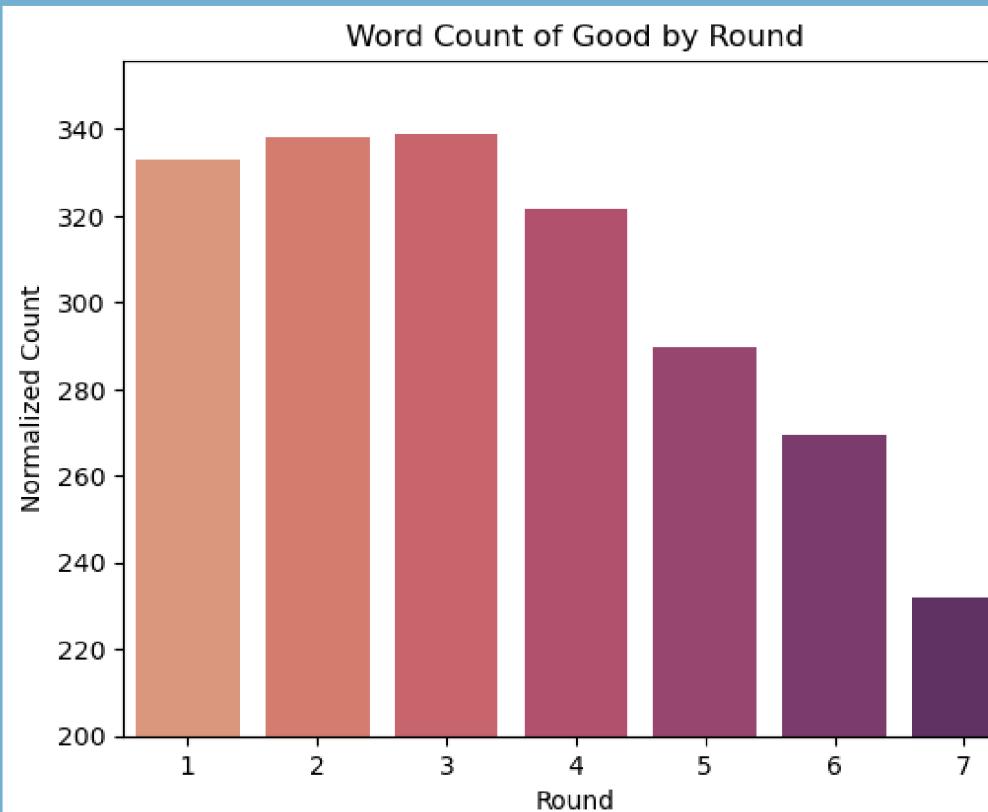
Word Count of Starter by Round



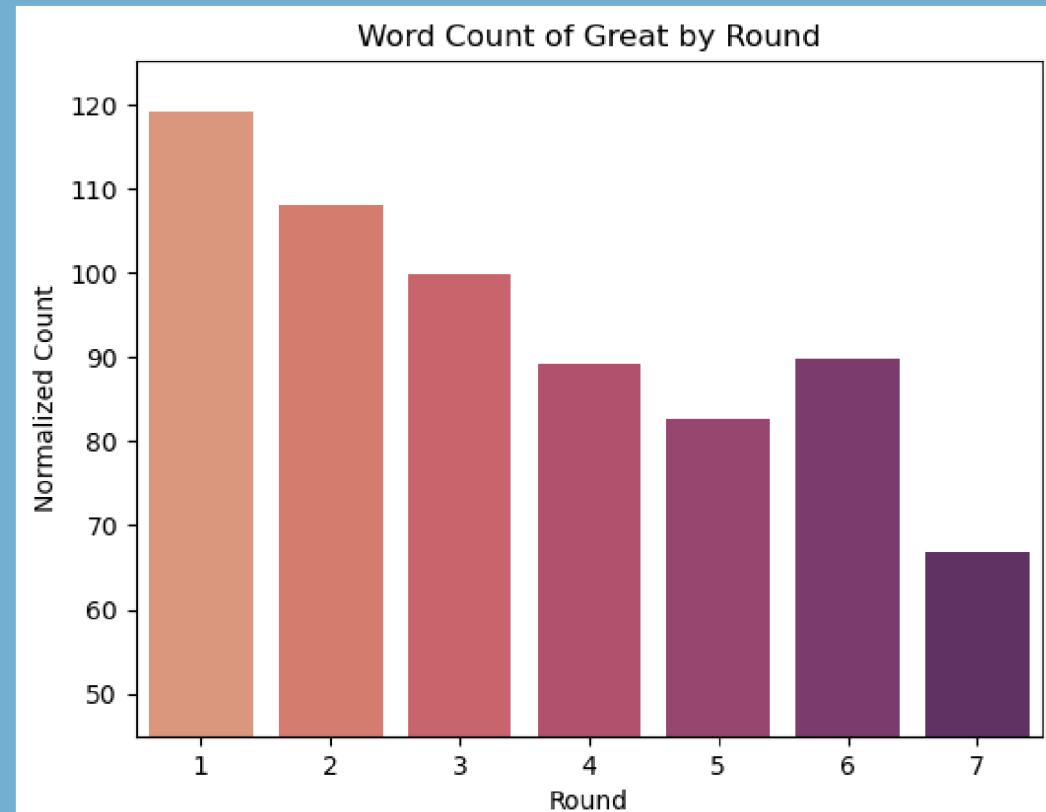
Word Count of Average by Round



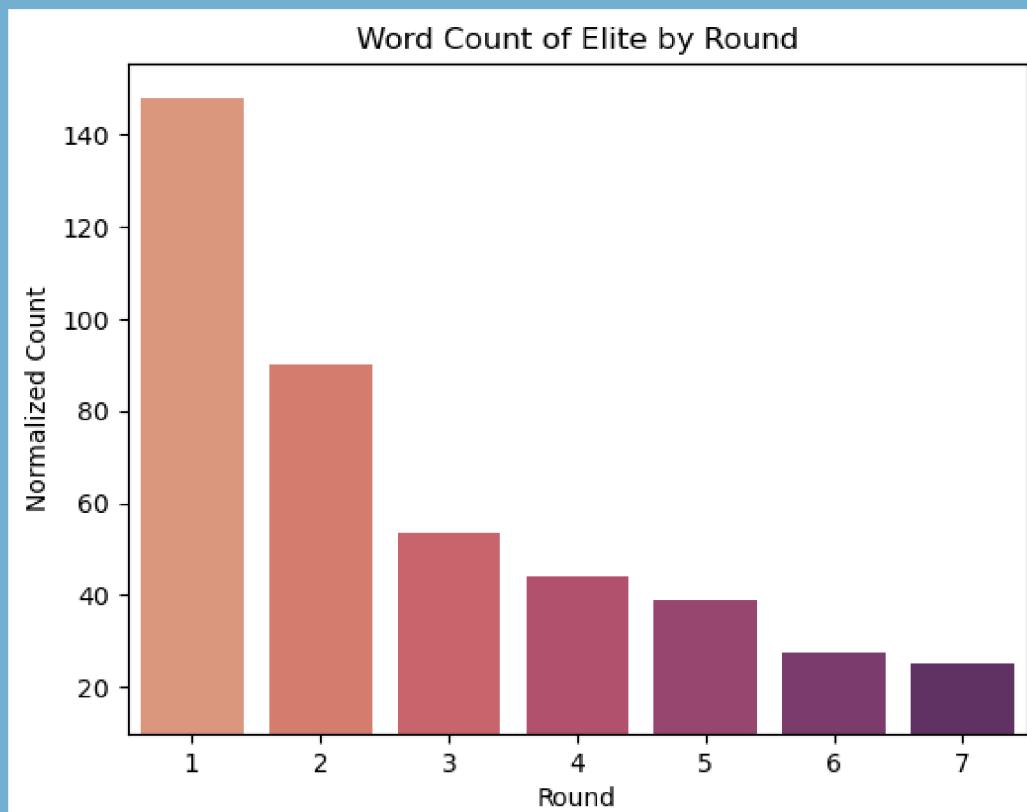
Word Count of Good by Round



Word Count of Great by Round



Word Count of Elite by Round





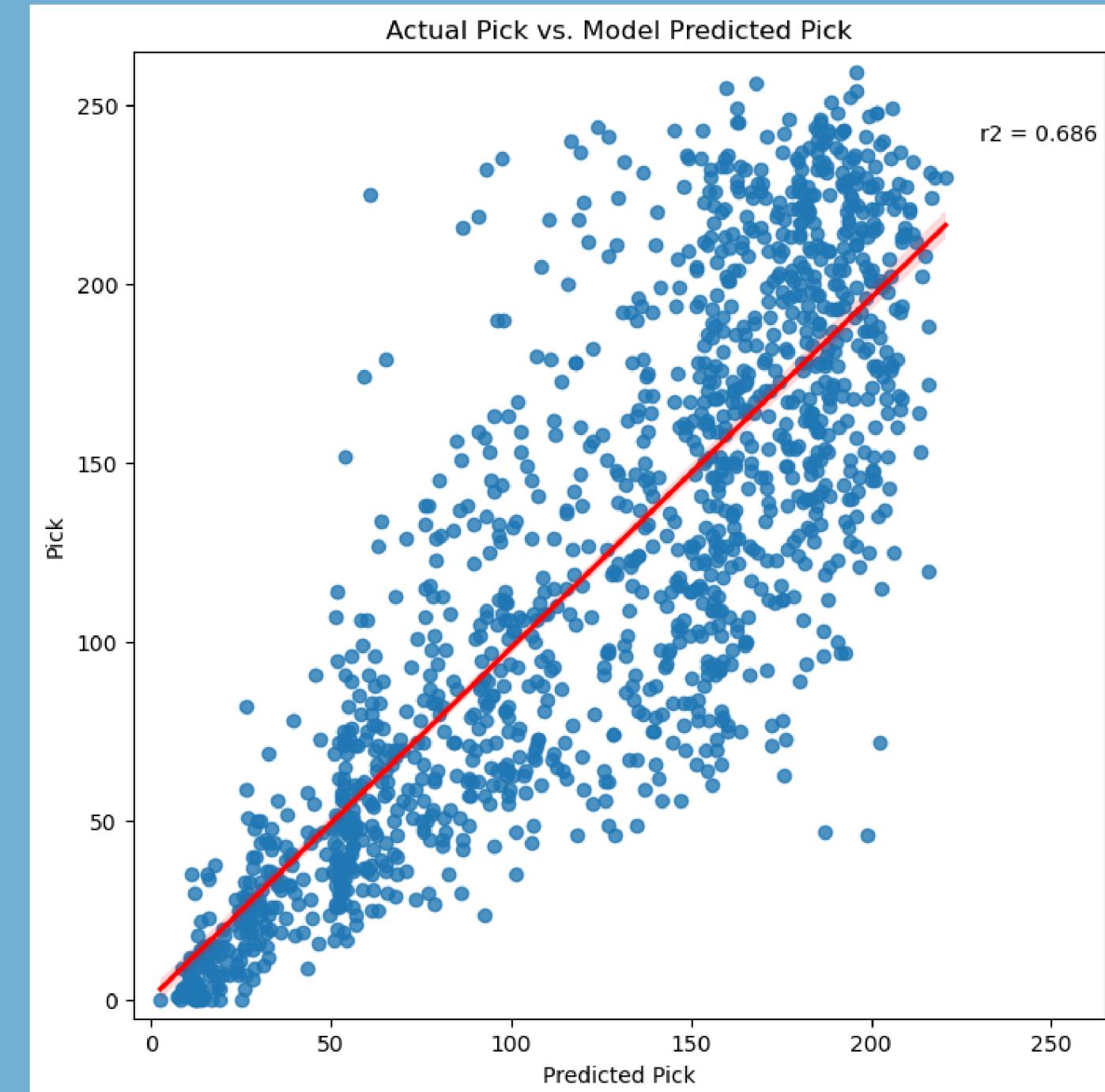
MODELS

	Linear Regression	KNN	Random Forest TFIDF	Random Forest CountVectorizer	Decision Tree	ADABoost	XGBoost	GradientBoost CountVectorizer	GradientBoost TFIDF
R ²	0.55	0.24	0.68	0.69	0.42	0.67	0.68	0.69	0.69
Mean Squared Error	2264	3809	1608	1583	2932	1654	1629	1579	1587



BEST MODEL

- Best Model: Gradient Boost (Using CountVectorizer)
- Parameters:
 - Learning Rate: .05
 - Max Depth: 4
 - Min Sample Leaf: 2
 - Min Sample Split: 2
 - N_Estimators: 92
 - RandomOverSampler





TOP FEATURES

Gradient Boost

Column	Weights
Overall Rank	0.80
Grade	0.09
Report Length	0.008
Forty Yard	0.005
Broad Jump	0.003
Lack	0.002

Random Forest

Column	Weights
Overall Rank	0.60
Grade	0.10
Report Length	0.007
Forty Yard	0.003
Broad Jump	0.003
Lack	0.004



CONCLUSIONS

- Grade and Overall Rank are the most correlated
- Lack, Average, Good, Great, Elite, and Starter among the most significant words in scouting reports
- It is easier to predict picks in the early rounds (especially round 1)
- Get more numerical data
 - Advanced stats
 - Non Primary Stats
 - More bio information
 - Spend more time feature engineering
- Medical Data
- Factor in team needs / which team is picking



QUESTIONS?

- **Josh Palgon**
- linkedin.com/in/josh-palgon/
- <https://github.com/jpalgon>
- jopalgon@gmail.com